HDT - Proposal / Vision

Definition of „top level perspectives"

The HDT project is a collection of Hadoop related plugins for the Eclipse plattform. It offers developers easy access to hadoop configuration, developement and operations.

Outlook and Wishes for an integrated Hadoop Toolbox

We want to deliver a toolbox for hadoop related developement tasks for Map Reduce jobs, InputFormats and OutputFormats. Developement of connectors to HDFS, HBase or Zookeeper should also be supported. In order to extend Hive and Pig, the developement of UDF and I/O connectors for Hive and Pig should be supported. Finally we want to be able to create Oozie workflows and submit them to a cluster.

Therefore we have to distinguish the different views / perspectives:

a) Developer-Perspective
b) Data-Connector-Perspectice
c) Cluster-Connector-Perspective
d) Operator-Perspective
e) Analysis-Perspecitve

Developer-Perspective

The source code (for Hadoop projects, not the pugins) is maintained here.
Wizards allow the creation of new Mapper, Reducer, Driver, UDF, etc. based on existing templates.
MRUnit is integrated and Jobs can be tested in LocalJobRunner mode or even sent to a connected TEST-cluster.
Artefacts are packaged and can be submitted to the central repository for deployment.

Cluster-Connector-Perspective

The Data-Connector Perspective allows the direct access to several services like HDFS, JobTracker, Zookeeper, HBase. The first level will be just browsing in internal webbrowser windows, without typing adresses and ports. All necessary information is retrieved from cluster configuration and mapped to a tree-view from there services can be selected. Icons in the tree-view indicate the status of the service.

Data-Connector-Perspectice

The data-connector perspective is used to connect to metastores like the Hive metastore or the HCatalog system.

Operator-Perspective

This perspective allows to start jobs, to store job templates and to analyse performance based on the analysis of the job.xml file (see Starfish for details).

Analysis-Perspecitve

The Analysis perspective is used for service orcestration. Ozzie Workflows can be created and submitted. Results can be retrieved from the cluster and used for BIRT based reports.